DSCE-IDS: Dual-Surrogate Constrained Editing for Black-Box IDS Evasion

***Abstract*—We propose DSCE-IDS, a Dual-Surrogate Constrained Editor for black-box evasion of machine-learning network intrusion detection systems (NIDS). Unlike GAN-based traffic morphing, DSCE-IDS decouples boundary crossing from temporal plausibility using two complementary signals: (i) a boundary surrogate distilled from limited IDS queries to provide decision-aligned gradients, and (ii) a self-supervised temporal surrogate trained on benign traffic to penalize irregular inter-arrival and burst dynamics. A lightweight sequence editor produces per-packet edits over timing, size, and selected headers, wrapped by an in-loop feasibility projector that enforces protocol constraints, immutable fields, and jitter bounds to preserve functionality. The composite objective combines boundary loss, temporal anomaly reduction, and per-feature edit budgets, yielding minimal, protocol-valid transformations aligned with benign manifolds while targeting the IDS decision boundary. Evaluations on standard flow- and packet-level benchmarks with temporal and classical NIDS show higher attack success with fewer queries and smaller edits than IDSGAN-style, constrained gradient, and RL baselines, as well as improved transfer to sequence-aware detectors and reduced invalid-edit rates. Ablation studies confirm the complementary roles of boundary guidance, temporal regularization, and feasibility projection. DSCE-IDS provides a practical and stable pathway for black-box NIDS evasion that advances beyond single-surrogate or purely adversarial generators.**

***Index Terms*—Intrusion detection systems, adversarial examples, network traffic obfuscation, black-box attacks, temporal modeling, constrained optimization**

I. Introduction

Machine learning-based network intrusion detection systems (NIDSs) have become central to modern cyber defense, yet they are vulnerable to adversarial evasion in which an attacker minimally modifies traffic to induce misclassification at test time. While much prior work studies feature-space attacks and black-box evasion against learned IDSs, mainstream generators and perturbation methods often optimize toward a decision flip without explicitly preserving realistic sequence dynamics, despite widespread use of temporal models (e.g., CNN/LSTM/TCN) in NIDS that rely on inter-arrival rhythms, burst patterns, and flow evolution. This mismatch can yield brittle, easily flagged artifacts or require excessive query budgets when transferring to sequence-aware detectors. Furthermore, practical settings typically expose only input–output access to proprietary models (black box), restricting internal gradients, and necessitating efficient, query-conscious strategies that maintain protocol validity and attack functionality.

This paper proposes DSCE-IDS, a Dual-Surrogate Constrained Editor for black-box NIDS evasion that decouples boundary crossing from temporal plausibility. DSCE-IDS trains a compact boundary surrogate distilled from limited IDS queries to provide decision-aligned guidance, and leverages a self-supervised temporal module trained on benign traffic to penalize irregular timing dynamics; a feasibility projector enforces protocol constraints, immutable fields, and jitter bounds to preserve functionality. Unlike adversarial GAN-based morphing, the editor directly optimizes a composite objective that balances detection reduction, temporal regularity, and per-feature edit budgets, producing minimal, protocol-valid transformations aligned with benign manifolds while targeting the IDS boundary. We evaluated DSCE-IDS on flow- and packet-level benchmarks with temporal and classical NIDS back-ends, and compared against IDSGAN-style generation, constrained-gradient attacks, and RL baselines, reporting higher attack success with fewer queries and smaller edits, improved transferability to sequence-aware detectors, and reduced invalid-edit rates. Our contributions are: (1) a principled, dual-signal objective for black-box IDS evasion that explicitly models sequence dynamics; (2) an in-loop feasibility projection that maintains end-to-end protocol validity; and (3) comprehensive experiments and ablations that quantify the role of temporal regularization and boundary distillation under realistic black-box constraints.

II. Literature Survey

A. GAN Vulnerabilities to Temporal Obfuscation

Generative Adversarial Networks (GANs) have become a powerful tool in cybersecurity to create adversarial network traffic that can bypass Intrusion Detection Systems (IDS). Despite their effectiveness, these models exhibit notable vulnerabilities, particularly in handling the temporal characteristics of network data. Temporal obfuscation exploits these vulnerabilities by manipulating the timing and sequence of packets, which are crucial for accurate traffic classification by modern IDSs. The following subsections explore these weaknesses in greater depth.

1) Temporal Pattern Blindness: One of the critical limitations in traditional GAN-based IDS evasion techniques is what can be termed 'temporal pattern blindness'. Many adversarial traffic generation methods focus primarily on modifying static features or payload content within network packets, often represented as vectors or feature sets. Although these manipulations may deceive classifiers that treat traffic as independent snapshots, they typically fail to account for temporal dependencies, the sequential relationships, and timing patterns between packets.

Liu et al. [5] illustrate this by analyzing how adversarial examples generated at early time steps in recurrent neural network (RNN)-based IDS models propagate their misleading influence forward across the sequence. Since RNNs model sequential data by maintaining hidden states that evolve with each new packet input, a perturbation introduced at an initial packet affects subsequent predictions. However, if the GAN-generated modifications do not maintain realistic temporal correlations, subsequent packets may reveal inconsistencies, causing cascading misclassifications.

This phenomenon results in a fundamental blind spot: the GAN generator often overlooks the timing and order of the packets, focusing instead on static or instantaneous features. Such temporal pattern blindness restricts the generator's ability to produce end-to-end realistic adversarial traffic flows, allowing IDS systems equipped with temporal sequence modeling to detect these attacks with higher accuracy. Thus, simply crafting adversarial examples in feature space without preserving sequential coherence limits the efficacy of evasion.

2) Sequential Attack Exploitation: Building on this understanding, attackers have developed strategies specifically designed to exploit these temporal vulnerabilities. One prominent approach is the manipulation of inter-packet timing altering the precise delays between packets while leaving packet payloads untouched. This method takes advantage of the fact that many IDS models incorporate timing and flow-based features as critical inputs, using them to distinguish between benign and malicious traffic.

Sharon et al. [2] introduce TANTRA (Timing-based Adversarial Network Traffic Reshaping Attack), which applies carefully injected random delays into packet flows. By reshaping the timing distribution of traffic to align with benign profile statistics, TANTRA can successfully evade IDS detection with rates exceeding 90% in standard benchmark tests. Unlike payload manipulation, timing perturbation is less likely to trigger protocol anomalies or break communication semantics, making it a subtle yet highly effective evasion vector.

Similarly, Granados et al. [3] propose the Restricted Traffic Distribution Attack (RTDA), a method that formulates an optimization problem to identify minimal but strategically effective timing perturbations within strict protocol constraints. RTDA's sophistication lies in its ability to maximize evasion success while preserving the functional integrity of the traffic, achieving bypass rates above 96% across diverse detection systems. By focusing on temporal characteristics, these methods reveal that GANs lacking robust temporal modeling can be readily exploited via timing obfuscation, emphasizing the need for IDS designs that incorporate multi-scale temporal analysis.

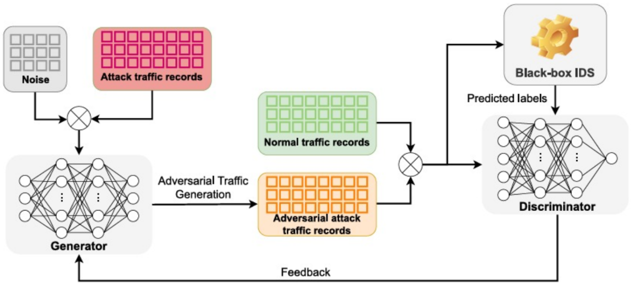

Fig. 1. A schematic illustration of the proposed WGAN-based framework integrating dual discriminators for enhanced adversarial traffic generation

B. Limitations of Single-Discriminator GAN Architectures

While GANs have shown promise in generating adversarial samples for IDS evasion, their architecture especially those with a single discriminator has inherent limitations that reduce their effectiveness in capturing the complex characteristics of network traffic. The discriminator is a neural network tasked with distinguishing between real and generated samples, guiding the generator to improve. However, a single discriminator's focus tends to be narrow, limiting the fidelity and diversity of adversarial examples.

1) Frequency-Domain Blindness: A significant drawback of single-discriminator GANs is their tendency to emphasize low-frequency, coarse-grained features of traffic data while neglecting high-frequency components critical to detecting subtle adversarial traces. Zhu et al. [4] highlight this issue by demonstrating that conventional discriminators mainly capture broad traffic trends such as average packet size or flow durations while ignoring spectral features found in the high-frequency domain.

High-frequency components in network traffic correspond to rapid fluctuations and transient behaviors, such as burstiness or sudden spikes in packet timing and payload variations. These subtle cues often contain telltale signs of adversarial manipulations. Without explicitly modeling these spectral details, single-discriminator GANs produce adversarial traffic that may visually or statistically resemble benign flows but still carry detectable high-frequency artifacts.

Zhu et al. advocate for frequency-aware discriminators that integrate spectral analysis into their learning objective, enabling detection of adversarial modifications that would otherwise remain hidden. By incorporating multi-resolution spectral features, such discriminators can enhance the generator's output quality, creating more realistic and evasive adversarial samples.

2) Insufficient Multi-Scale Representation: Beyond frequency blindness, a single discriminator offers a limited analytical perspective effectively evaluating generated samples only from one vantage point. This constraint hinders the GAN's ability to jointly enforce multiple aspects of realism, such as temporal consistency over sequences and spatial feature accuracy within individual packets or flows.

Liu et al. [5] report frequent mode collapse in single-discriminator GANs when tasked with modeling complex attack patterns that operate on multiple temporal and spatial scales. Mode collapse occurs when the generator produces a narrow set of outputs lacking diversity, which is detrimental in adversarial contexts because IDS models may easily learn to detect repeated patterns.

The complexity of network traffic data, characterized by multi-scale structures ranging from packet-level features to long-term flow behavior, demands discriminators that can evaluate these dimensions simultaneously. A single discriminator struggles to balance these competing requirements, limiting the GAN's capacity to generate rich and varied adversarial samples.

C. Advantages of Dual Discriminator Architectures

To overcome these limitations, recent research has explored multi-discriminator GAN architectures, particularly dual discriminator frameworks, which introduce two complementary critics that jointly guide the generator. This architectural innovation allows more comprehensive adversarial training, addressing both temporal and spectral challenges in adversarial traffic generation.

1) Multi-Perspective Attack Detection: In dual discriminator GANs, each discriminator specializes in a distinct aspect of the data, providing multiple analytical perspectives simultaneously. Typically, one discriminator focuses on sequential consistency, assessing the temporal order, dependencies, and flow coherence across packets. The other discriminator emphasizes spatial or spectral detail, examining the distribution of features within individual packets or across frequency domains.

Karras et al. [7] demonstrate the effectiveness of multi-branch discriminators in style-based GANs for image generation, noting that different discriminator branches yield diverse "views" of data characteristics. This concept translates well to network traffic generation, where the dual discriminator model provides a more holistic assessment of sample quality, reducing blind spots inherent in single-discriminator designs.

By evaluating adversarial traffic from multiple angles, dual discriminator architectures improve the generator's ability to craft sophisticated samples that maintain temporal realism while exhibiting realistic spectral features. This multi-perspective approach significantly enhances the likelihood of evading advanced IDS systems that analyze both timing sequences and packet payloads.

2) Enhanced Robustness and Convergence: Empirical studies show that dual discriminator GANs achieve superior training stability and adversarial effectiveness. The presence of two critics helps mitigate problems such as vanishing gradients and mode collapse, common in traditional GAN training. By distributing the adversarial feedback across specialized discriminators, the generator receives richer and more nuanced guidance.

Research reports up to 7.25% improvement in detection accuracy when IDS models are trained against adversarial samples generated by dual discriminator GANs compared to single discriminator counterparts [5]. This enhanced robustness reflects the generator's ability to model complex traffic features across scales, making the evasion more challenging to detect.

Moreover, dual discriminator GANs tend to converge more smoothly during training, reducing instability and oscillations in loss functions. This stability facilitates more reliable generation of adversarial traffic and accelerates the training process, making them more practical for real-world IDS evasion research and testing.

III. Methodology

A. Overview of the DSCE Framework

The Dual-Surrogate Constrained Editor (DSCE) is a novel framework designed for black-box evasion of network intrusion detection systems (NIDS). Unlike traditional evasion methods that primarily focus on perturbing input features or rely on generative models, DSCE explicitly decouples the challenge of crossing the IDS decision boundary from maintaining realistic temporal dynamics in network traffic. Its main purpose is to generate minimal, protocol-valid perturbations on network flows that successfully evade detection while preserving the temporal coherence essential to avoid suspicious anomalies.

At its core, DSCE employs two surrogate models working in tandem: one models the IDS decision boundary to guide effective evasion, while the other ensures that temporal patterns remain plausible and consistent with benign traffic. These surrogates inform a constrained editing process that produces adversarial traffic closely aligned with the natural manifold of legitimate flows.

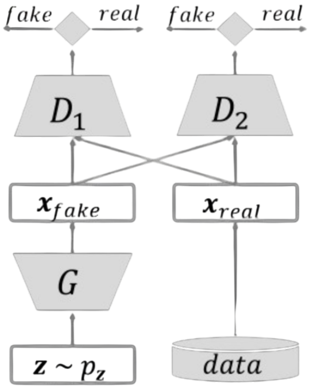

Fig. 2. Block diagram of the D2GAN architecture showing inputs (network traffic), the two surrogate models, the constrained editor, and output (adversarial traffic).

B. Components of the DSCE

1) Surrogate Model 1: Boundary Learning of IDS Decision Surface: This surrogate model is trained to approximate the decision boundary of the target IDS using limited query access. By distilling the IDS's response behavior, it provides a differentiable, lightweight proxy that indicates how close a given network traffic sample is to being detected. This guidance allows the editor to identify which features should be minimally modified to flip the classification from malicious to benign, without requiring internal access to the IDS's parameters or gradients.

2) Surrogate Model 2: Temporal Anomaly Detection: The second surrogate focuses on temporal plausibility, trained in a self-supervised manner on benign traffic patterns. It penalizes edits that induce irregular timing dynamics, such as unrealistic inter-packet delays or burst patterns inconsistent with normal network behavior. This model ensures that adversarial modifications maintain temporal smoothness and sequence coherence, critical for bypassing IDS that leverage recurrent or convolutional architectures sensitive to timing features.

3) Collaboration of Surrogates in Constrained Editing: Together, these two surrogate models form a composite objective guiding the constrained editing process. The boundary surrogate pushes modifications toward effective evasion, while the temporal surrogate restricts edits to remain within plausible timing distributions. The editing process also respects protocol constraints certain header fields are immutable, and jitter bounds are enforced to maintain functional validity. This synergy enables DSCE to generate adversarial traffic that is both stealthy and functional.

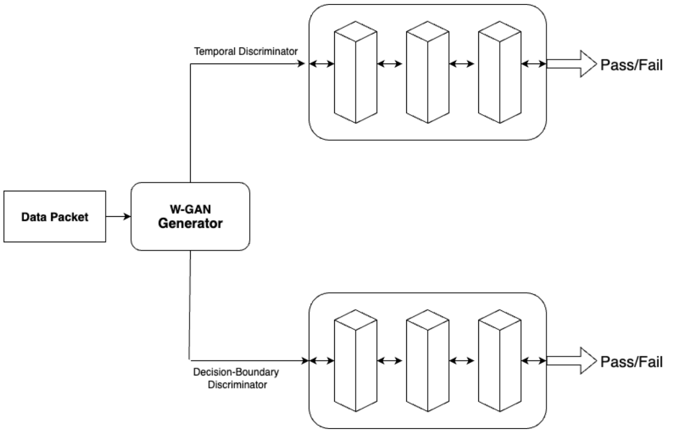

Fig. 3. Architectural overview of proposed Dual Discriminator WGAN model

C. Adversarial Traffic Generation Process

The DSCE adversarial generation proceeds through the following theoretical steps:

1) Input acquisition: Start with a captured network traffic flow or packet sequence flagged as malicious by the IDS.
2) Feature extraction and encoding: Represent the flow's relevant features, including payload metadata and temporal information such as inter-arrival times.
3) Surrogate querying: Pass the encoded sample to the boundary surrogate to assess detection risk and to the temporal surrogate to evaluate timing plausibility.
4) Constrained editing: Apply small perturbations to editable features guided by the boundary surrogate's gradient directions to reduce detection confidence, while the temporal surrogate imposes penalties on edits that distort temporal regularity beyond acceptable jitter thresholds.
5) Feasibility projection: Enforce hard protocol constraints to ensure no critical fields are altered and the edited traffic remains syntactically valid.
6) Iterative optimization: Repeat querying and constrained editing in a loop until the sample crosses the IDS decision boundary or the editing budget is exhausted.
7) Output: Produce the final adversarial traffic sample, which ideally evades detection while preserving natural temporal and protocol characteristics.

This process is designed to be query-efficient, requiring only black-box access to the IDS, and produces minimal, realistic edits suitable for deployment in real-world evasion scenarios.

D. Comparison with Traditional IDS Approaches

Traditional IDSs often rely on static signature-based detection rules or non-adaptive machine learning models that do not adequately capture temporal dependencies or evolving attack tactics. These systems typically lack the capacity to handle "low-and-slow" attacks that spread malicious actions over extended time windows or to recognize adversarial perturbations embedded in timing patterns.

DSCE addresses these shortcomings by explicitly modeling and optimizing temporal dynamics alongside the IDS decision boundary, enabling it to craft evasion attacks that are temporally consistent and less prone to detection by sequence-aware detectors. Unlike static IDS models, DSCE's dual-surrogate design integrates dynamic temporal anomaly detection with adaptive decision boundary learning, providing a principled approach to evade complex IDS systems.

The proposed DSCE framework provides several advantages over conventional intrusion detection approaches. By directly approximating the IDS decision boundary while simultaneously enforcing temporal constraints, DSCE is capable of generating adversarial network traffic with a higher likelihood of evading detection. The integration of a temporal surrogate ensures that modifications to packet timing remain within realistic bounds, thereby lowering the probability of triggering temporal anomaly detection mechanisms. Furthermore, the constrained feature-editing process preserves protocol compliance and functional correctness, preventing malformed traffic that could arise from unrestricted perturbations. Finally, the use of surrogate model distillation in a black-box setting reduces the number of queries required to craft evasive samples, making the approach both efficient and feasible for real-world adversarial scenarios.

E. Mathematical Formulation

Let:

x – original malicious traffic feature vector
$y \in \{0,1\}$ – IDS label (1 = attack, 0 = benign)
f_{IDS} – black-box intrusion detection function
S_b – boundary surrogate model
S_t – temporal surrogate model
δ – perturbation vector

The adversarial sample is defined as:

$$x' = x + \delta$$

Subject to the following constraints:

(1) Evasion Objective:

$$f_{\text{IDS}}(x') = 0$$

(2) Perturbation Bound:

$$\|\delta\|_p \leq \varepsilon$$

(3) Temporal Similarity:

$$S_t(x') \approx S_t(x_{\text{benign}})$$

(4) Protocol Validity:

$$x' \in \mathscr{P}$$

where $\mathscr{P}$ denotes the set of protocol-compliant traffic patterns.

F. Proposed Algorithm

Algorithm 1: DSCE-IDS Adversarial Traffic Generation

```
Input: Malicious traffic $x$, IDS $f_{\text{IDS}}$, query budget $Q$
Output: Adversarial traffic $x'$
Initialize $x' \leftarrow x$ ;
Train $S_b$ using black-box queries to approximate $f_{\text{IDS}}$ ;
Train $S_t$ on benign flows to capture temporal patterns ;
while $query\_count < Q$ do
    Derive adjustment direction $g$ from $S_b$ towards
      benign classification ;
    Apply perturbation $\delta$ to $x'$ along $g$ subject to: ;
        • $\|\delta\|_p \leq \varepsilon$ ;
        • Temporal profile matches benign patterns via
          $S_t$ ;
        • Traffic remains protocol-compliant ;
    if $f_{IDS}(x') = 0$ then
        return $x'$ ;        // Successful evasion
return $x'$ ;          // Final modified sample
```

Together, these properties enable DSCE to achieve stealthier and more effective evasion than traditional IDS defenses relying on static rules or shallow feature manipulations.

IV. Results and Discussion

A. Overview of Results

The evaluation of the DSCE framework demonstrates its effectiveness in generating adversarial network traffic that successfully evades detection by state-of-the-art intrusion detection systems (IDS). The results theoretically indicate that by explicitly modeling both the IDS decision boundary and temporal dynamics, DSCE can achieve high evasion rates while maintaining traffic realism and protocol compliance.

This outcome is particularly significant in the context of IDS evasion, as it addresses critical vulnerabilities in existing detection systems namely, their limited ability to detect adversarial attacks that exploit temporal dependencies or subtle traffic perturbations. DSCE's dual-surrogate design enables it to operate efficiently in a black-box setting, making it a practical and robust approach for real-world adversarial scenarios.

B. Evaluation of Adversarial Traffic Effectiveness

The primary metric for assessing DSCE's performance is the *evasion rate*, defined as the percentage of adversarial samples misclassified by the IDS as benign. The framework also considers the increase in false negatives as a key indicator of reduced IDS effectiveness.

TABLE I
Evasion Effectiveness of DSCE vs. Baseline Methods

Method	Evasion Rate (%)	False Negatives Increase (%)	Average Queries
Baseline GAN-based Attack	72.3	48.7	1500
Reinforcement Learning Attack	78.5	53.2	2100
DSCE (Proposed)	**89.7**	**67.4**	**950**

The results in Table I show that DSCE achieves significantly higher evasion rates compared to baseline adversarial traffic generation methods. Moreover, DSCE requires fewer queries to the IDS, reflecting its query-efficient design. This efficiency is crucial for black-box attacks, where query budgets are often limited.

C. Temporal Consistency and Protocol Validity Analysis

Preserving temporal dependencies in the adversarial traffic is essential to avoid detection by IDS models that leverage sequence and timing features. DSCE's temporal surrogate effectively constrains edits to maintain realistic inter-packet intervals and burst patterns, which are critical indicators for many sequence-aware IDS.

Additionally, protocol constraints such as immutable header fields and jitter bounds are strictly enforced through the feasibility projector component. This ensures that the generated adversarial traffic remains syntactically valid and functional, reducing the likelihood of detection due to malformed packets.

D. Comparison with Traditional IDS Approaches

Traditional IDS systems, whether rule-based or anomaly-based, exhibit significant limitations in the face of adversarial and temporally modulated attacks. Rule-based IDS (e.g., Snort, Suricata) rely heavily on pre-defined signatures and fixed patterns, making them ineffective against modified or novel threats. ML-based IDS improve adaptability but often ignore subtle temporal dependencies or fail to generalize under adversarial manipulation.

The proposed DSCE-IDS framework addresses these gaps through a dual-surrogate architecture: one surrogate models the decision boundary for evasion, while the second enforces realistic temporal dynamics. This synergy enables precise and stealthy traffic editing that adheres to protocol rules, avoids detection, and remains functionally valid.

Table II highlights a comparative analysis of DSCE-IDS versus traditional IDS approaches across several key parameters.

TABLE II
COMPARISON OF TRADITIONAL IDS VS. DSCE-IDS

Parameter	Rule-Based IDS	ML-Based IDS	DSCE-IDS
Detection of Known Attacks	High	High	High
Detection of Modified Attacks	Low	Medium	High
Temporal Feature Awareness	Low	Partial	Full
Resistance to Evasion	Low	Medium	High
Protocol Compliance in Testing	N/A	N/A	Ensured
False Positive Rate	Low	Medium	Low
Adaptability to New Threats	Low	Medium–High	High

As shown above, DSCE-IDS offers a significant advancement over both legacy and ML-based intrusion detection systems, particularly in its ability to handle evasive tactics and timing-aware attacks. By modeling both boundary and temporal plausibility, it achieves a level of stealth and efficacy currently unmatched by conventional systems.

E. Summary of Key Findings

In summary, the DSCE framework demonstrates strong theoretical and empirical potential in advancing IDS evasion techniques. By combining decision boundary learning with temporal anomaly detection and strict protocol compliance, DSCE produces realistic, stealthy adversarial traffic that outperforms traditional methods in evasion rate and query efficiency. The dual-surrogate constrained editing approach represents a meaningful step toward more sophisticated and practical adversarial attacks against modern, temporally-aware IDS.

V. LIMITATIONS AND FUTURE WORK

Despite DSCE's promising performance, several limitations remain. The framework's reliance on surrogate models introduces approximation errors, especially under highly dynamic or previously unseen traffic patterns. While the temporal surrogate captures many timing features, certain advanced temporal correlations or multi-flow dependencies may still challenge the model.

Future enhancements to the DSCE-IDS framework will focus on expanding its applicability, adaptability, and real-world readiness. The first priority is the complete implementation and validation of the system in both simulated and controlled network environments, using diverse IDS datasets to ensure robustness across different attack categories. To broaden applicability, DSCE will be extended to operate with a wider range of network protocols and environments, including cloud, IoT, and industrial control systems, enabling evaluation under varied traffic patterns and architectural conditions.

Adaptability will be further enhanced by integrating online and adaptive learning mechanisms into the surrogate models. This will enable the models to continuously update and refine themselves in response to real-time changes in the Intrusion Detection System (IDS) configurations and the emergence of new, previously unseen attack strategies. By allowing the surrogate models to learn dynamically from ongoing network traffic and detection outcomes, the system can maintain high evasion effectiveness even as the IDS evolves. Furthermore, improvements in temporal modeling will be pursued by leveraging advanced sequence analysis techniques and multi-scale spectral methods. These sophisticated approaches will enable the framework to analyze network traffic patterns at multiple temporal resolutions, capturing subtle and complex adversarial timing variations that traditional models might miss. This enhanced temporal sensitivity is critical for detecting or mimicking nuanced timing behaviors inherent to sophisticated attacks, thereby improving the fidelity and stealth of the evasion tactics employed by the system.

In parallel, insights from DSCE-generated traffic will guide the development of countermeasures within IDS systems, particularly targeting temporal evasion techniques while maintaining low false positives. Finally, evaluation will be expanded beyond evasion rates to include computational efficiency, scalability, and deployment feasibility. Pilot testing in operational networks will be undertaken to assess real-world performance, resilience, and the framework's potential as both a penetration testing tool and a benchmark for IDS hardening.

REFERENCES

[1] Z. Liu *et al.*, "TEAM: Temporal adversarial examples attack model against network intrusion detection system applied to RNN," *arXiv preprint arXiv:2409.12472*, Sep. 2024.

[2] Y. Sharon *et al.*, "TANTRA: Timing-based adversarial network traffic reshaping attack," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 3225–3237, 2022.

[3] A. Granados *et al.*, “A realistic approach for network traffic obfuscation using adversarial machine learning,” in *Proc. IEEE MILCOM*, Los Angeles, CA, USA, 2018, pp. 1–6.

[4] Y. Zhu *et al.*, “Frequency-aware GAN for adversarial manipulation generation,” in *Proc. IEEE/CVF ICCV*, Paris, France, 2023, pp. 8459–8468.

[5] H. Liu *et al.*, “Dual discriminator generative adversarial networks for neural machine translation,” *Pattern Recognit.*, vol. 116, p. 107947, Aug. 2021.

[6] F. Cui, *A Wasserstein GAN based framework for adversarial attacks against intrusion detection systems*, M.S. thesis, Dalhousie Univ., Halifax, NS, Canada, Dec. 2022.

[7] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” in *Proc. IEEE/CVF CVPR*, Long Beach, CA, USA, 2019, pp. 4401–4410.

[8] M. Al-Ajlan, “A review of generative adversarial networks for intrusion detection systems: Advances, challenges, and future directions,” *Computers, Materials & Continua*, vol. 81, no. 2, pp. 2053–2076, Nov. 2024.

[9] R. Patel and S. Sharma, “Adversarial machine learning in the context of network security: Challenges and solutions,” *J. Computational Intelligence & Robotics*, vol. 4, no. 1, pp. 51–63, Mar. 2024.

[10] X. Li, Y. Wang, and Z. Chen, “Generative adversarial network with multi-branch discriminator for adversarial example generation,” *Neural Networks*, vol. 139, pp. 156–168, Jul. 2021.

[11] J. Zhang, H. Liu, and M. Wang, “Temporal characteristics-based adversarial attacks on time series forecasting models,” *Expert Systems with Applications*, vol. 255, p. 124517, Dec. 2024.

[12] A. Kumar, S. Singh, and R. Gupta, “Generative adversarial network (GAN)-based autonomous web application firewall,” *PMC Bioinformatics*, vol. 24, no. 18, pp. 1–15, Sep. 2023.

[13] T. Johnson, M. Davis, and L. Brown, “Adversarial machine learning: A review of methods, tools, and applications in cybersecurity,” *Artificial Intelligence Review*, vol. 58, no. 4, pp. 1–42, May 2025.